ChIP-seq: welcome to the new frontier

Elaine R Mardis

Next-generation sequencing technology combines with chromatin immunoprecipitation to provide a genome-wide look at transcription-factor binding.

Next-generation sequencing technologies, capable of producing tens of millions of sequence reads during each instrument run, are quickly being applied in a myriad of creative ways to answer genome-wide questions. In this issue, Robertson and colleagues describe such an application, comparing chromatin immunoprecipitation (ChIP) of the Stat1 transcription factor using a next-generation sequencing platform and a conventional microarray-based platform[1].

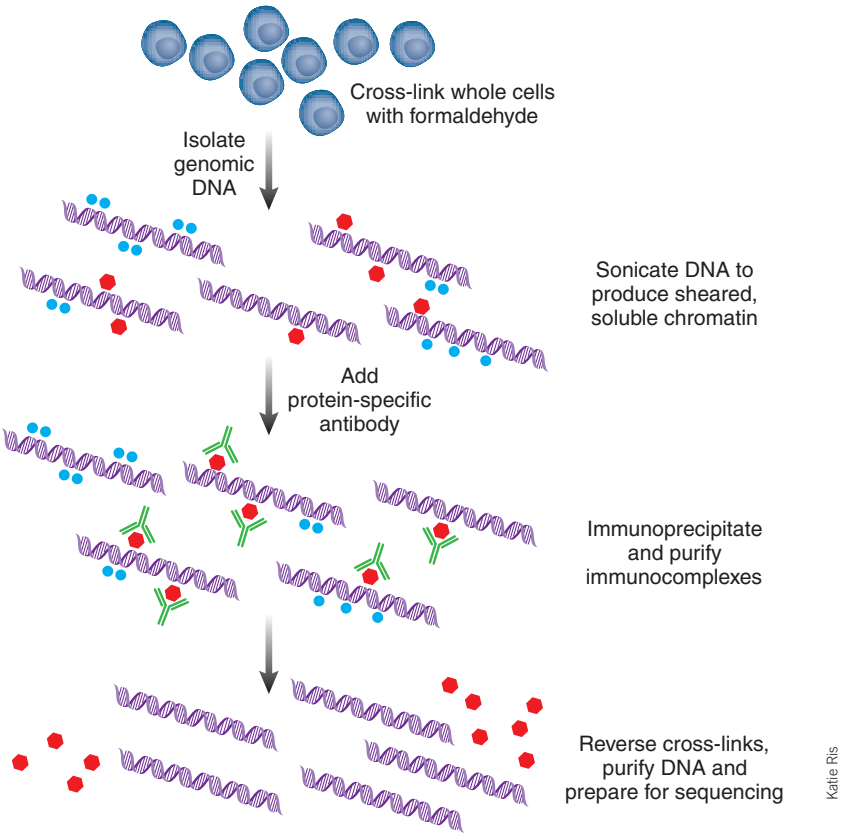

Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

This report provides an elegant example of the power of next-generation sequencing platforms to expand what once was a focused assay to a genome-wide scope. In the process, our ability to characterize and understand phenomena such as alterations in transcription-factor binding in response to environmental stimuli can be evaluated for the entire genome in a single experiment. Hence, the ramifications for the pace of biological inquiry and the functional annotation of genomes are profound.

ChIP, first described by Varshavsky and colleagues[2] as a method to study protein-DNA interactions, comprises three basic steps. First, covalent cross-links between proteins and DNA are formed, typically by treating cells with formaldehyde or another chemical reagent. In the second step, an antibody specific to the protein of interest is used to selectively coimmunoprecipitate the protein-bound DNA fragments that were covalently cross-linked. Finally, the immunoprecipitated protein-DNA links are reversed and the recovered DNA is assayed to determine the sequences bound by that protein (**Fig. 1**). Because random protein-DNA cross-linking can occur, and nonspecific DNA can be pulled down in the immunoprecipitation step, the ChIP-selected DNA is typically compared to a mock sample of DNA collected without antibody addition during the immunoprecipitation step. Typically, these two DNA populations are differentially labeled and compared by hybridization to a genomic microarray ('ChIP-chip'), as initially reported by Ren and colleagues[3] in yeast.

Although ChIP-chip approaches have greatly expanded our understanding of genome-wide protein-DNA associations, the substitution of next-generation sequencing technology to analyze the DNA fragments released after ChIP ('ChIP-seq') has distinct advantages over microarray hybridization. As shown in Robertson *et al.*[1], the Solexa sequencing technology[4] provided short read length sequences of ~30 base pairs that were ideal for characterizing ChIP-derived fragments. The

Elaine R. Mardis is at the Washington University School of Medicine, St. Louis, Missouri 63108, USA.
e-mail: emardis@watson.wustl.edu

researchers mapped the resulting sequences back to the reference genome, whereby the most frequently sequenced fragments formed peaks at specific genomic regions. They then analyzed sequences under these peaks by comparison known Stat1 binding site sequences and locations, and for their proximity to genes. They also compared the results to previous Stat1 ChIP-chip data. Their comparison of Stat1 binding locations in human HeLa S3 cells stimulated by interferon gamma versus unstimulated cells showed that stimulated cells provided evidence of fourfold more Stat1-bound sites, and that specific sites bound in these interferon gamma–stimulated cells correlated well with expectations from previous studies[5,6]. Peaks for both unstimulated and stimulated cells showed the highest density at approximately 100 base pairs upstream of the transcriptional start sites of nearby genes.

In summary, ChIP-seq offers important advantages over ChIP-chip, including lower cost, minimal hands-on processing and a requirement for fewer replicate experiments as well as less input material. Moreover, the Stat1 experimental ChIP-seq data have a high degree of similarity to results obtained by ChIP-chip for the same type of experiment, with >64% of peaks in shared genomic regions. Because the data are sequence reads, ChIP-seq offers a rapid analysis pipeline (as long as a high-quality genome sequence is available for read mapping) as well as the potential to detect mutations in binding-site sequences, which may directly support any observed changes in protein binding and gene regulation.

This report, and a similar study by Johnson and colleagues[7] that defined genome-wide binding sites of the neuron-restrictive silencer factor (NRSF), provide initial evidence that next-generation sequencing platforms like Solexa are being cleverly coupled with previously focused techniques to produce genome-wide views of protein-binding phenomena. Variations on this theme have also provided data that delineate changes in histone methylation and thus give insights into chromatin packaging of the human genome[8,9].

In the near future these sequencing platforms are likely to identify new microRNAs and provide quantitative tag-based gene expression data, for example. These approaches will be combined to complete the picture of biological phenomena such as embryonic stem cell development and cancer progression. Their net impact will be to rapidly enhance the functional annotation of genomes, well beyond our present understanding. Improved annotation will in turn dramatically improve our ability to interpret sequence-level variations that concomitantly are being discovered by resequencing of genomes using next-generation technologies.

COMPETING INTERESTS STATEMENT
The author declares no competing financial interests.

1. Robertson, G. *et al. Nat. Methods* **4,** 651–657 (2007).
2. Solomon, M.J., Larsen, P.L. & Varshavsky, A. *Cell* **53**, 937–947 (1988).
3. Ren, B. *et al. Science* **290**, 2306–2309 (2000).
4. Bentley, D.R. *Curr. Opin. Genet. Dev.* **16**, 545–552 (2006).
5. Brierley, M.M. & Fish, E.N. *J. Interferon Cytokine Res.* **25**, 733–744 (2005).
6. Schroder, K. *et al. J. Leukoc. Biol.* **75**, 163–189 (2004).
7. Johnson, D.S. *et al. Science* **316**, 1497–1502 (2007).
8. Barski, A. *et al. Cell* **129**, 823–837 (2007).
9. Mikkelsen, T.S. *et al. Nature*; published online 1 July 2007.

Prion biology: the quest for the test

Adriano Aguzzi

The adaptation of the protein misfolding cyclic amplification assay (PMCA) to use recombinant hamster prion protein (PrP) as a substrate shows promise for both basic research applications and clinical diagnostic assays.

Prions, the elusive infectious particles responsible for transmissible spongiform encephalopathies (Creutzfeldt-Jakob disease in humans as well as various conditions in wild and captive animals), have been responsible for some of the most tragic disasters in the history of medicine[1]. For decades, biological materials derived from subclinically infected organisms have transferred the infection to others. Thus, the development of sensitive, fast and reliable assays represents an urgent and unmet medical need. In this issue of *Nature Methods*, Atarashi and colleagues from the Caughey laboratory introduce a technique that may be an interesting step toward that goal and may also help to understand the mechanisms of prion replication[2].

For a long time, blood products were not thought of as vectors of iatrogenic prion transmission. The bovine spongiform encephalopathy (BSE) epidemic has changed that perception. BSE has infected cows throughout the world, and as a result of eating tainted beef, some 160 people have contracted variant Creutzfeldt-Jakob disease (vCJD). An undetermined number of hitherto asymptomatic persons may have also contracted infection. Some of them have donated blood, resulting in at least four cases of blood-borne transmission of vCJD. Such transmission will be likely witnessed time and again in the coming decades[3].

PCR-based nucleic acid testing technologies have reached wondrous levels of sensitivity and speed: it is now possible to detect viruses in donated blood with sufficient sensitivity to fully exclude transmission. Yet prions do not contain nucleic acids and consist primarily of proteins—the sensitive detection of which lags well behind that of nucleic acids. These hurdles have become painfully evident when addressing the prion detection question.

According to the protein-only hypothesis, prions consist of PrP^{Sc}, a misfolded, structured aggregate of a cellular constituent called PrP^{C}. A very large body of evidence supports this hypothesis, and PrP-deficient mice that lack PrP^{C} are resistant to prion infection[4]. Because PrP^{C} is prevalent in body fluids and shares most epitopes with its aggregated counterpart, termed PrP^{Sc}, prions have proved resilient to immunodetection. In 1994, Caughey and colleagues discovered that PrP^{Sc} can catalyze the conversion of PrP^{C} into more PrP^{Sc}, but that process was inefficient and detectable only with considerable efforts[5].

Adriano Aguzzi is at the Institute of Neuropathology, CH-8091 Zürich, Switzerland.
e-mail: adriano.aguzzi@usz.ch